ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

RANKING FRAUD REVEALING SYSTEM FOR MOBILE APPs AND AVOIDANCE FROM USER'S RECOMMENDATION

Arun.D [1], Sumithradevi.C [2]

1. P.G. Student, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India
2. Asst.Professor, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India

Abstract: Ranking fraud in the mobile App business alludes to false or tricky exercises which have a motivation behind, knocking up the Apps in the fame list. To be sure, it turns out to be more incessant for App designers to utilize shady means, for example, expanding their Apps' business or posting imposter App evaluations, to confer positioning misrepresentation. While the significance of avoiding Ranking fraud has been generally perceived, there is constrained comprehension and examination here. This paper gives a holistic perspective of positioning misrepresentation and propose a Ranking fraud identification framework for mobile Apps. In particular, it is proposed to precisely find the mining so as to pose extortion the dynamic periods, to be specific driving sessions, of portable Apps. Such driving sessions can be utilized for identifying the neighborhood inconsistency rather than a worldwide abnormality of App rankings. Moreover, three sorts of proof s are explored, i.e., positioning based confirmations, modelling so as to rate based proofs and survey based proofs, Apps' positioning, rating and audit practices through factual theories tests. Also, a streamlining based conglomeration technique is explored to incorporate every one of the confirmations for misrepresentation discovery. At last, assessment of the proposed framework is done with certifiable App information gathered from the iOS App Store for quite a while period. In the investigations, this paper accepts the adequacy of the proposed framework, and demonstrates the identification's versatility calculation and also some consistency of positioning misrepresentation exercises.

KEYWORDS: Mobile Apps, ranking fraud detection, evidence aggregation, historical ranking records, rating and review

I INTRODUCTION

The quantity of mobile Apps has developed at an amazing rate in the course of recent years. For instances, the growth of apps were increased by 1.6 million at Apple's App store and Google Play. To fortify the improvement of portable Apps, numerous App stores dispatched every day App leaderboards, which exhibit the graph rankings of most prominent Apps. To be sure, the App leaderboard is a standout amongst the most essential courses for advancing mobile Apps. A higher rank on the leaderboard more often than not prompts countless and million dollars in income. In this way, App designers have a tendency to investigate different routes, for example, publicizing battles to advance their Apps keeping in mind the end goal to have their Apps positioned as high as could be expected under the circumstances in such App leaderboards. Be that as it may, as a late pattern, rather than depending on customary showcasing arrangements, shady App engineers resort to some fake intends to intentionally help their Apps and in the end control the diagram rankings on an App store. This is typically actualized by utilizing purported "bot homesteads" or "human water armed forces" to blow up the App downloads, evaluations and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

surveys in a brief while. For instance, an article from Venture Beat reported that, when an App was advanced with the assistance of positioning control, it could be impelled from number 1,800 to the main 25 in Apple's sans top leaderboard and more than 50,000-100,000 new clients could be gained inside of a few days. Truth be told, such positioning misrepresentation raises awesome worries to the portable App industry. For instance, Apple has cautioned of getting serious about App designers who confer positioning extortion in the Apple's App store.

There are some related works, for example, web positioning spam recognition online survey spam identification and portable App suggestion the issue of distinguishing positioning misrepresentation for mobile Apps is still under-investigated. To fill this essential void, in this paper, a system is build up for positioning misrepresentation discovery framework for portable Apps. Along this line, distinguishable essential difficulties are also considered. To begin with, positioning misrepresentation does not generally happen in the entire life cycle of an App, so recognition is done when the time when extortion happens. Such test can be viewed as recognizing the neighbourhood inconsistency rather than worldwide irregularity of mobile Apps. Second, because of the immense number of portable Apps, it is hard to physically mark positioning extortion for each App, so it is essential to have an adaptable approach to consequently recognize positioning misrepresentation without utilizing any benchmark data. At long last, because of the dynamic way of outline rankings, it is difficult to distinguish and affirm the confirmations connected to positioning misrepresentation, which rouses us to find some verifiable extortion examples of portable Apps as proofs. Surely, our watchful perception uncovers that mobile Apps are not generally positioned high in the leaderboard, but rather just in some driving occasions, which shape distinctive driving sessions. Note that it is presented both driving occasions and driving sessions in point of interest later. As such, positioning extortion more often than not happens in these driving sessions. In this way, distinguishing positioning misrepresentation of mobile Apps is really to identify positioning extortion inside of driving sessions of portable Apps. In particular, a model is proposed which is a basic yet compelling calculation to recognize the main sessions of each App in light of its verifiable ranking records. At that point, with the examination of Apps' positioning practices, the false Apps are found which frequently diverse positioning examples in every driving session contrasted and typical Apps. In this way, it describes some misrepresentation confirmations from Apps' chronicled positioning records, and build up three capacities to concentrate such positioning based extortion confirmations. In any case, the positioning based proofs can be influenced by App designers' notoriety and some honest to goodness advertising battles, for example, "restricted time rebate". Accordingly, it is not sufficient to just utilize positioning based proofs. In this manner, two sorts of extortion proofs are proposed taking into account Apps' evaluating and survey history, which mirror some irregularity designs from Apps' verifiable rating and audit records. Furthermore, we add to an unsupervised proof total system to incorporate these three sorts of confirmations for assessing the validity of driving sessions from portable Apps.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

II LITERATURE SURVEY

1.D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation" [1]:

D. M. Blei, A. Y. Ng, and M. I. Jordan, introduces a unique model called as Dirichlet allocation (LDA) a generative probabilistic model for collections of discrete data such as text amount. Basically it is a three level hierarchical Bayesian model in which each element of a group is demonstrated as a finite mixture over a fundamental set of topics. Each topic is demonstrated as an infinite mixture over fundamental set of topic probabilities. With the reference of text modelling, the topic probabilities provide an open representation of a document. An efficient approximation inference technique is presented based on various methods and an EM algorithm for empirical Bayes parameter estimation is also presented. The results are reported in document modelling, text classification and collaborative filtering, which compares to a collection of unigrams and probabilistic LSI model.

2.Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection systemices in city taxis"[2]:

Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou , illustrated that growth in the field of GPS tracking technology have allowed the users to install GPS tracking devices in taxies to gather huge amount of GPS traces under some time period. These traces by GPS offered an unparalleled opportunity to uncover taxi driving fraud traces and then fraud detection system is proposed which is able to identify taxi driving fraud. First, two sort of function are uncovered here i.e. travel route evidence and driving distance evidence. Even a third function is developed to combine the previous functions based on Dempster-Shafer theory. First identification of interesting locations is done from tremendous amount of taxi GPS logs and then parameter free method is proposed to extract the travel route evidences. Secondly, concept of route mark is developed to illustrate the driving path between locations and based on those mark, specific model is characterized for the distribution of driving distance and discover the driving distance evidences. And finally, taxi driving fraud detection system with a large scale real world taxi GPS logs.

3. T. L. Griffiths and M. Steyvers, "Rank aggregation via nuclear norm minimization" [3]:

T. L. Griffiths and M. Steyvers, introduces the process of rank aggregation which is interweave with the structure of skew-symmetric matrices. Recent development in the principles of matrix completion matrices is been applied and this idea gives rise to a new method for ranking a set of items. The core of this idea deals with the raking aggregation method which intimately describes a partially filled skew-symmetric matrix. The algorithm for matrix completion is raised to hold skew-symmetric data and use that to extract ranks for each item. This algorithm applies same strategy for both pairwise comparisons as well as for rating data. It becomes robust to noise and incomplete data as it is based on matrix completion.

4. A. Klementiev, D. Roth, and K. Small , "An unsupervised learning algorithm for rank aggregation"[6]:

A. Klementiev, D. Roth, and K. Small, describes the field of information retrieval, data mining, and natural language, many applications needs a ranking of instances which is not present in classification. Furthermore, a rank aggregation is a result of aggregating the results of the established

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

ranking models into a formalism and then result represents a novel unsupervised learning algorithm (ULARA)which gives a linear combination of individual ranking functions. These functions were developed based on the axiom of rewarding ordering agreement between the rankers.

5. A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models" [8]:

A. Klementiev, D. Roth, and K. Small, produces a model which has to integrate the set of rankings often deals with aggregating and it only comes up when a certain ranked data is developed. Even though the various heuristic and supervised learning approaches to rank aggregation, a requirement of domain knowledge and supervised ranked data exists. Therefore, to solve this issue, a framework is proposed for learning aggregate rankings without supervision. This framework is instantiated for the cases of permutations and combinations of top-k lists.

III SYSTEM ANALYSIS

EXISTING SYSTEM:

- In the literature, while there are some related work, such as web ranking spam detection, online review spam detection and mobile App recommendation, the problem of detecting ranking fraud for mobile Apps is still under-explored.
- Generally speaking, the related works of this study can be grouped into three categories.
- The first category is about web ranking spam detection.
- The second category is focused on detecting online review spam.
- Finally, the third category includes the studies on mobile App recommendation

DISADVANTAGES OF EXISTING SYSTEM:

- Although some of the existing approaches can be used for anomaly detection from historical rating and review records, they are not able to extract fraud evidences for a given time period (i.e., leading session).
- Cannot able to detect ranking fraud happened in Apps' historical leading sessions
- There is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud.

PROPOSED SYSTEM:

- We first propose a simple yet effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we characterize some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences.
- We further propose two types of fraud evidences based on Apps' rating and review history, which reflect some anomaly patterns from Apps' historical rating and review records.
- In Ranking Based Evidences, by analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors in a

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase.

- In Rating Based Evidences, specifically, after an App has been published, it can be rated by any user who downloaded it. Indeed, user rating is one of the most important features of App advertisement. An App which has higher rating may attract more users to download and can also be ranked higher in the leaderboard. Thus, rating manipulation is also an important perspective of ranking fraud.
- In Review Based Evidences, besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspective of App ranking fraud.

ADVANTAGES OF PROPOSED SYSTEM:

- The proposed framework is scalable and can be extended with other domain generated evidences for ranking fraud detection.
- Experimental results show the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities.
- To the best of our knowledge, there is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud. Thus, we develop four intuitive baselines and invite five human evaluators to validate the effectiveness of our approach Evidence Aggregation based Ranking Fraud Detection (EA-RFD).
- Detect Fraud ranking in daily App leader boards.
- Avoid ranking manipulation.

IV. SYSTEM ARCHITECTURE

In recent years, mobile app has been growing tremendously while boosting more than 400,000 applications like Apple app store and Google Android market. This tremendous growth of mobile App has made it difficult to user for finding unique and trusted patterns of Application in crowded App stores. Thus to solve this important issue, existing marketing executives precisely use the App download history and ratings by the users to recommend the mobile applications which is totally trusted. Identifying ranking fraud is actually to identify ranking fraud of mobile apps within such leading sessions. In this paper, an useful algorithm is used to discover the leading sessions based on the historical records and with the help of analysis of those records, it is proved that deceptive apps usually have different ranking patterns in each leading sessions as compared to the normal apps. Therefore it is illustrated from those ranking records that some fraud is taking place in mobile app market and to restrict those frauds, three main evidences are developed to detect such fraud. As only ranking based evidences does not seems to be much sufficient to detect the fraud of mobile app, based on apps rating and review history some fraud evidences were discovered which showed anomaly patterns by those history. Specifically, an

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

unsupervised evidence aggregation method is also proposed here to integrate all such types of evidences for evaluating the trustworthiness of leading sessions. And finally, the proposed system is estimated with real world app data gathered from the Apple's app store for time consuming period, i.e., more than two years. The results of these experiments showed an effectiveness of proposed approach in fig 1.

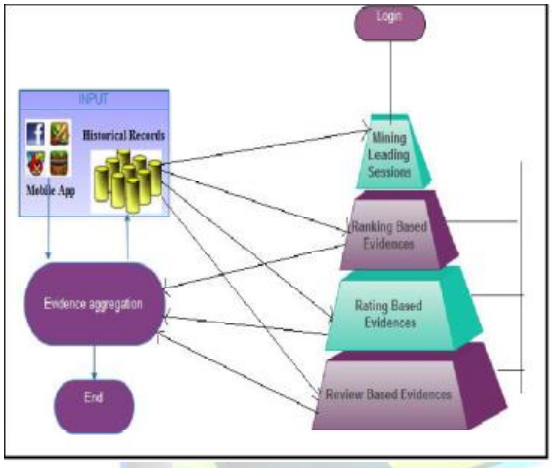

Fig.1. A Framework of our ranking fraud detection system for Mobile App.

A. MODULES DESCRIPTION.

a) Module 1: Leading events

Given a positioning limit K*∈[1, K] a main occasion e of App a contains a period range also, relating rankings of a, Note that positioning edge K * is applied which is normally littler than K here on the grounds that K may be huge (e.g., more than 1,000), and the positioning records past K * (e.g., 300) are not exceptionally helpful for recognizing the positioning controls. Moreover, it is finding that a few Apps have a few nearby driving even which are near one another and structure a main session.

b) Module 2: Leading Sessions

Instinctively, mainly the leading sessions of mobile app signify the period of popularity, and so these leading sessions will comprise of ranking manipulation only. Hence, the issue of identifying ranking fraud is to identify deceptive leading sessions. Along with the main task is to extract the leading sessions of a mobile App from its historical ranking records.

c) Module 3: Identifying the leading sessions for mobile apps

Basically, mining leading sessions has two types of steps concerning with mobile fraud apps.Firstly, from the Apps historical ranking records, discovery of leading events is done and then secondly merging of adjacent leading events is done which appeared for constructing leading sessions.Certainly, some specific algorithm is demonstrated from the pseudo code of mining sessions of given mobile App and that algorithm is able to identify the certain leading events and sessions by scanning historical records one by one.

d) Module 4: Identifying evidences for ranking fraud detection

1. Ranking based evidences:

It concludes that leading session comprises of various leading events. Hence by analysis of basic behaviour of leading events for finding fraud evidences and also for the app historical ranking records, it is been observed that a specific ranking pattern is always satisfied by app ranking behaviour in a leading event.

2. Rating based evidences:

Previous ranking based evidences are useful for detection purpose but it is not sufficient. Resolving the problem of “restrict time reduction”, identification of fraud evidences is planned due to app historical rating records. As we know that rating is been done after downloading it by the user, and if the rating is high in leaderboard

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 12, March 2016

considerably that is attracted by most of the mobile app users. Spontaneously, the ratings during the leading session gives rise to the anomaly pattern which happens during rating fraud. These historical records can be used for developing rating based evidences.

3. Review based evidences:

We are familiar with the review which contains some textual comments as reviews by app user and before downloading or using the app user mostly prefer to refer the reviews given by most of the users. Therefore, although due to some previous works on review spam detection [14], [19], [21], there still issue on locating the local anomaly of reviews in leading sessions. So based on apps review behaviors, fraud evidences are used to detect the ranking fraud in Mobile app

V. CONCLUSION AND FUTURE WORK

This paper introduces a system which is built up and it is actually a positioning extortion discovery framework for mobile Apps. In particular, initially it is demonstrated that positioning misrepresentation happened in driving sessions and gave a system to digging driving sessions for each App from its chronicled positioning records. At that point, it is recognized that positioning based confirmations, rating based proofs and survey based confirmations are used for identifying positioning extortion. In addition, a unique model is proposed which is an improvement based total system to incorporate every one of the proofs for assessing the validity of driving sessions from portable Apps. A novel point of view of this methodology is that every one of the proofs can be displayed by measurable theory test, in this way it is anything but difficult to be reached out with different confirmations from space information to distinguish positioning misrepresentation. At last, the proposed framework is accepted with broad examinations on certifiable App information gathered from the Apple's App store. Exploratory results demonstrated the adequacy of the proposed methodology. Later on, to concentrate more viable misrepresentation confirms and dissect the idle relationship among rating, survey and rankings is panned. In addition, amplification of positioning misrepresentation location approach is performed with other portable App related administrations, for example, mobile Apps suggestion, for improving client experience.

REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., pp. 993–1022, 2003.

[2] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection system," in Proc. IEEE 11th Int. Conf. Data Mining, 2011, pp. 181–190.

[3] D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 60–68.

[4] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.

[5] G. Heinrich, Parameter estimation for text analysis, " Univ. Leipzig, Leipzig, Germany, Tech. Rep., http://faculty.cs.byu.edu/~ringger/CS601R/papers/Heinrich-GibbsLDA.pdf, 2008.

[6] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int.Conf. Web Search Data Mining, 2008, pp. 219–230.

[7] A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank

aggregation," in Proc. 18th Eur. Conf. Mach. Learn., 2007, pp. 616–623.

[8] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models," in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 472–479.

[9] A. Klementiev, D. Roth, K. Small, and I. Titov, "Unsupervised rank aggregation with domain-specific expertise," in Proc. 21st Int. Joint Conf. Artif. Intell., 2009, pp. 1101–1106.

[10] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proc.19thACMInt. Conf. Inform. Knowl. Manage., 2010, pp. 939–948.

[11] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li, "Supervised rank aggregation," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 481–490.

[12] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 632–640.

[13] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf. World Wide Web, 2006, pp. 83–92.

[14] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[15] K. Shi and K. Ali, "Getjar mobile application recommendations with very sparse datasets," in Proc. 18th ACM SIGKDD Int. Conf. Knowl.Discovery Data Mining, 2012, pp. 204–212.

[16] N. Spirin and J. Han, "Survey on web spam detection: Principles and algorithms," SIGKDD Explor. Newslett., vol. 13, no. 2, pp. 50–64, May 2012.